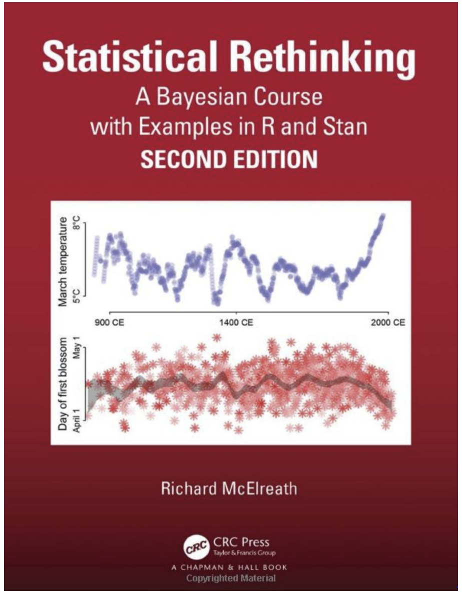
CSSE Lab Session 5

Testing Models

Johannes Härtel
(johannes.hartel@vub.be)



[McElreath20]

The major source for this lecture.

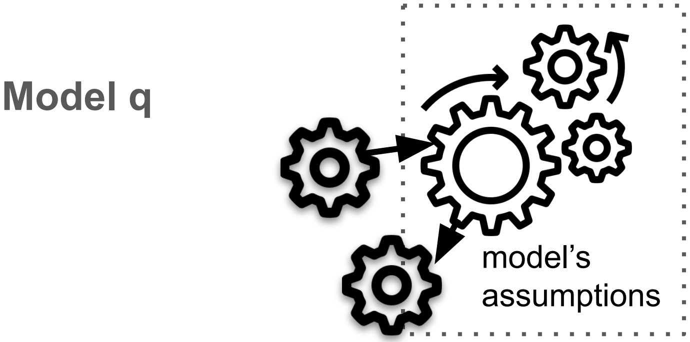
The problem of testing models

The model comparison story

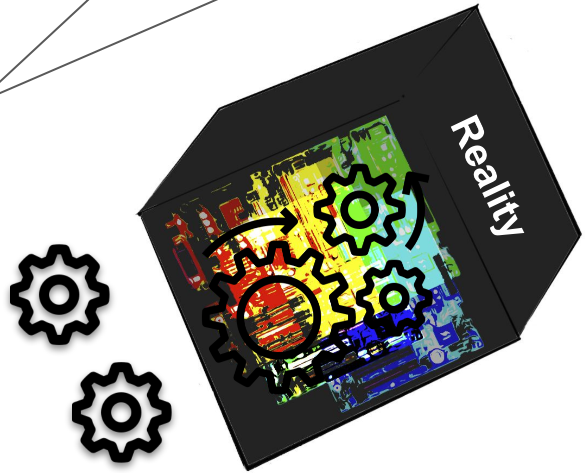
Understanding (modeling) reality, that is a **black box**.

The model comparison story

Modeling the reality.

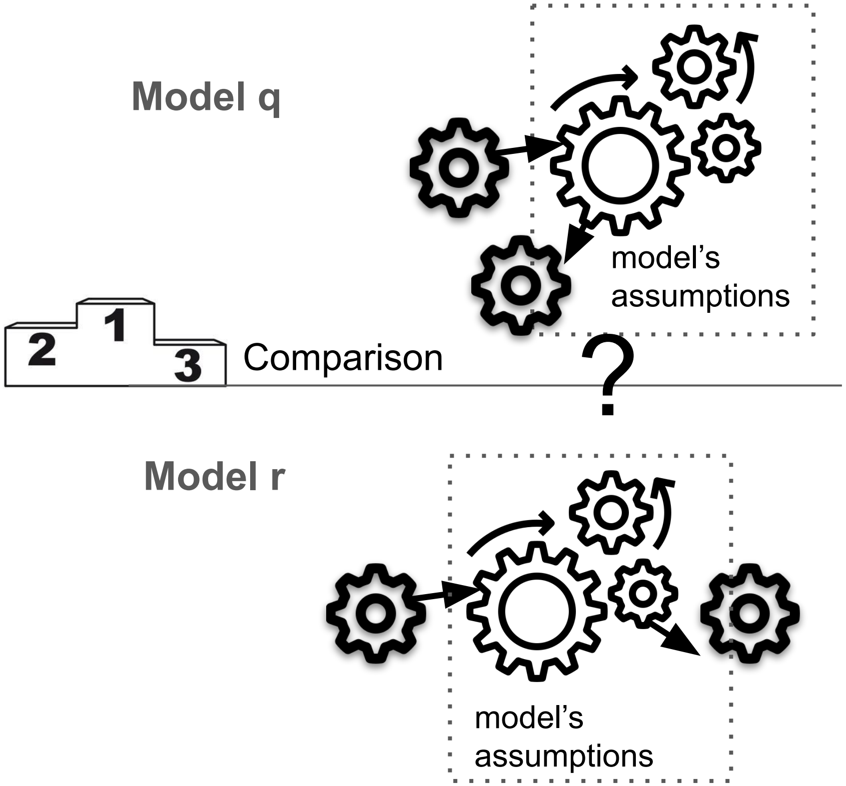


How do we know that the model meets reality?



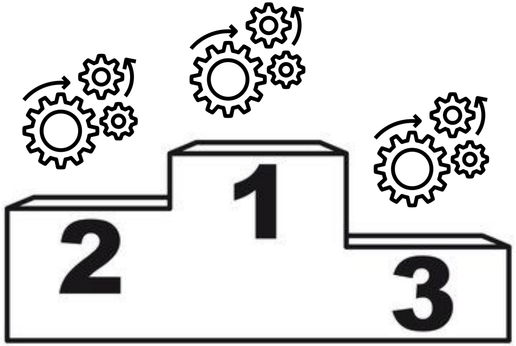
The model comparison story

The model (or classifier) that **fits the data better**, should be **closer to the reality**.



We can compare the "**fit**" of both models without knowing what is inside **the black box.**

How to **measure** which model is ‘better’ while doing the comparison?



For this lecture, we will look at a linear regression model.

(You find it as the "linear component" in the logistic regression.)

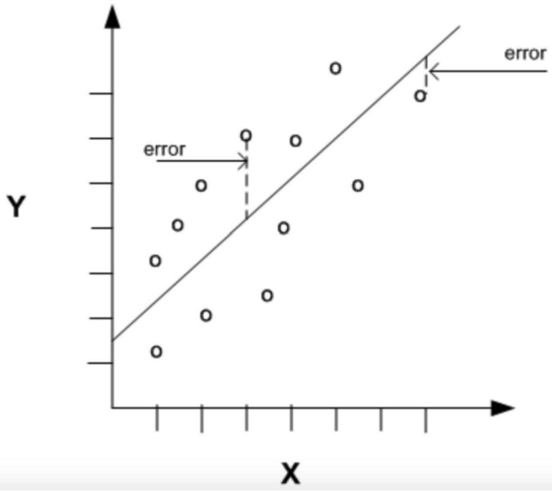
recap …

evaluating linear regression

All · Images · Videos · News · Books · More · Tools

About 202.000.000 results (0,36 seconds)

R Square/Adjusted R Square : This is a first measure of regression model especially we, everybody, do during evaluation because it is easy to interpret score between 0 to 1. If we see good score like close to 1, then we assume that model is good fit. 30 Jul 2020

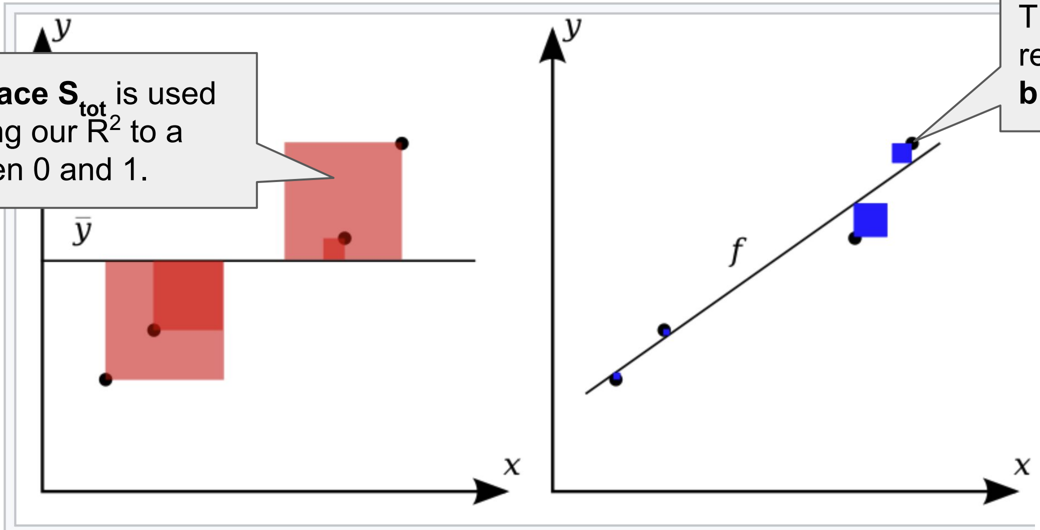


As a side note, there are **many other measures**.

Measuring the fit of a linear regression

For linear regression, one often uses R^2 to measure the fit.

The **red surface S_{tot}** is used for normalizing our R^2 to a range between 0 and 1.



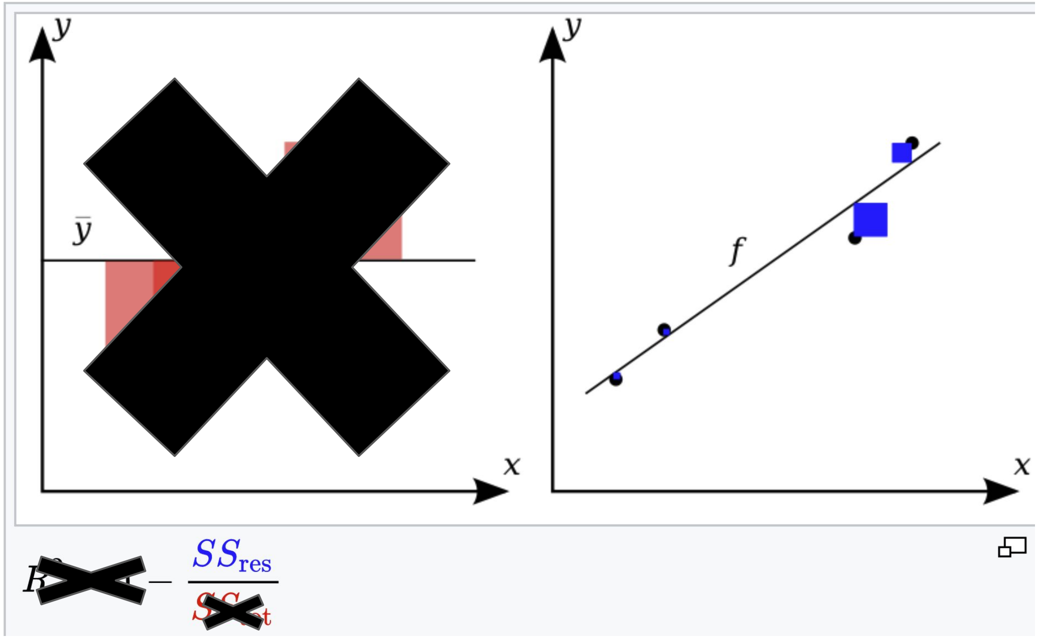
The closer the black dot comes to the regression line (the error), the smaller the **blue surface** (our square error), i.e., $\mathbf{SS_{res}}$.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Source: https://en.wikipedia.org/wiki/Coefficient_of_determination

Occasionally, we can ignore the red surface.

Then we have some aggregate of the square error (e.g., mean or sum of square error). It surfaced when comparing on the same data (if red stays the same).



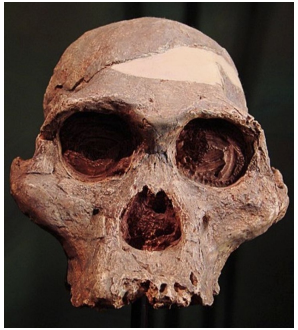
Source:
https://en.wikipedia.org/wiki/Coefficient_of_determination

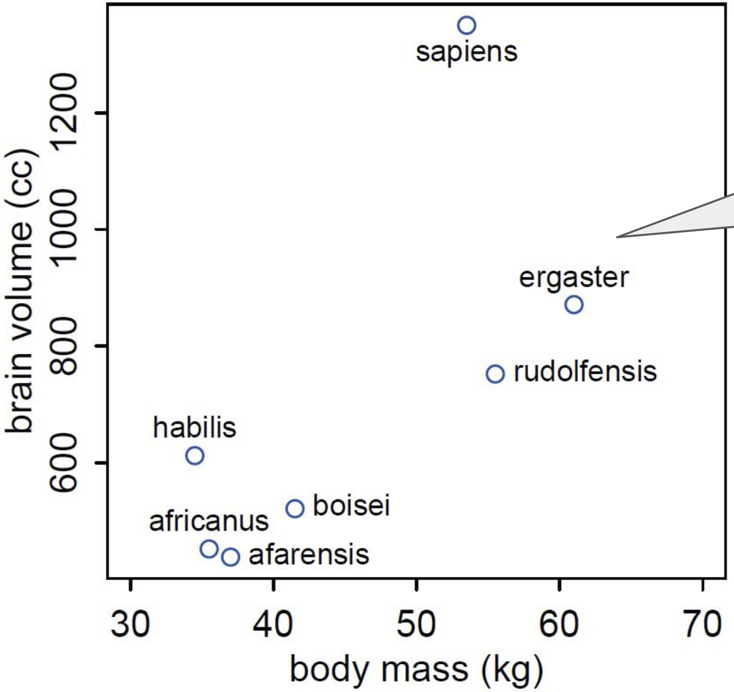
The pitfall

(Why do we need cross-validation?)

Example: Examining the relation between body mass and brain volume using polynomial (linear) regression.

We don't know relation between both.



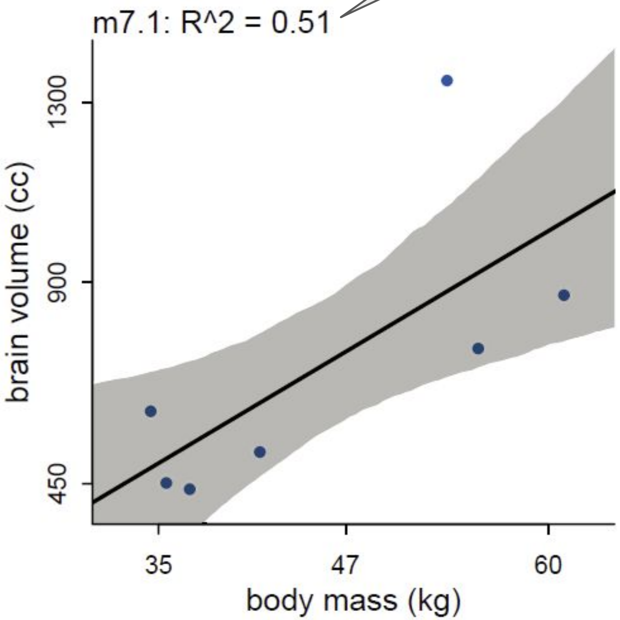


The **observed variable** *body mass* and *brain volume on x and y-axis*.

(Copy [McElreath20])

Example: Polynomial regression (degree 1)

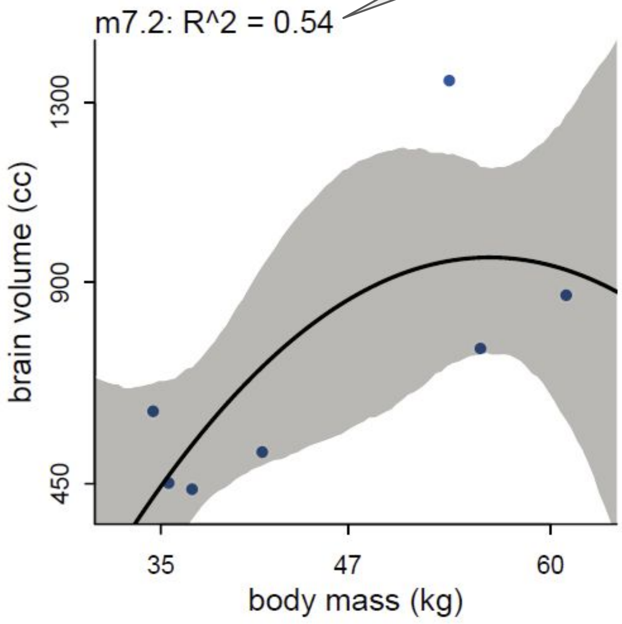
Fit R^2



(Copy [McElreath20])

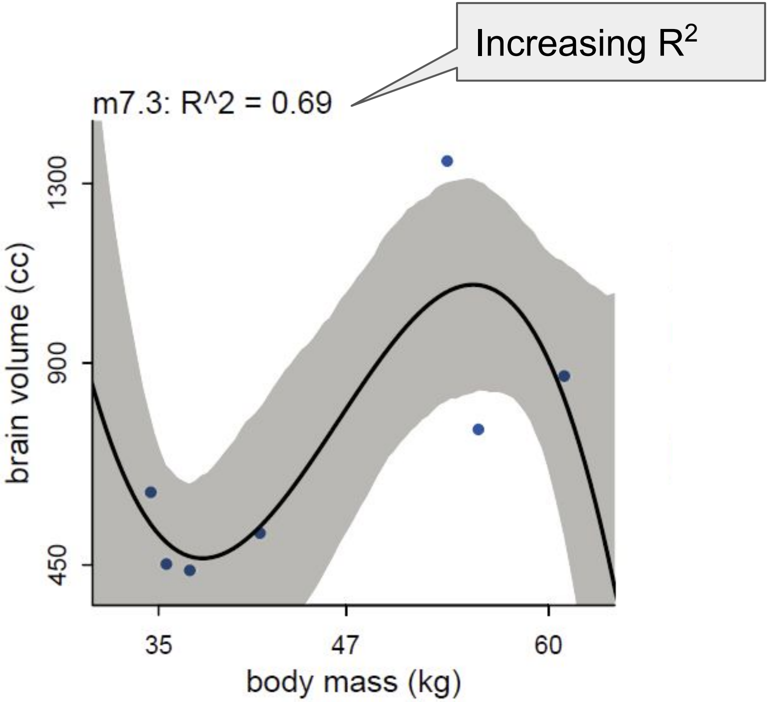
Example: Polynomial regression (degree 2)

Increasing R^2



(Copy [McElreath20])

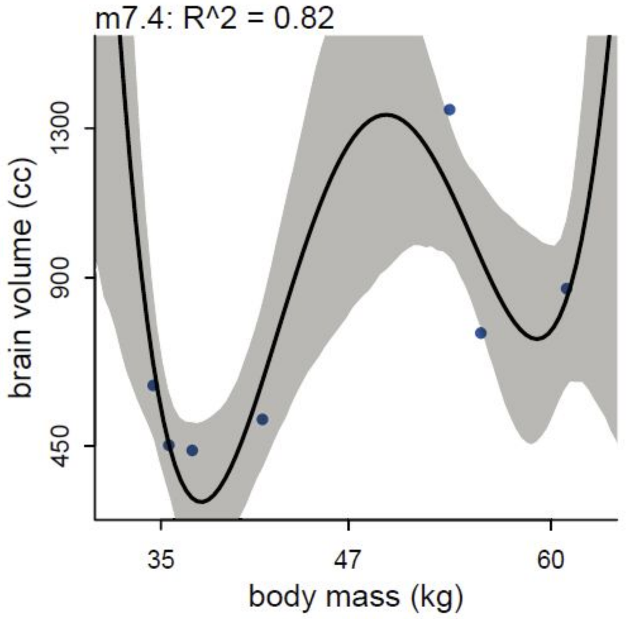
Example: Polynomial regression (degree 4)



(Copy [McElreath20])

Example: Polynomial regression (degree 4)

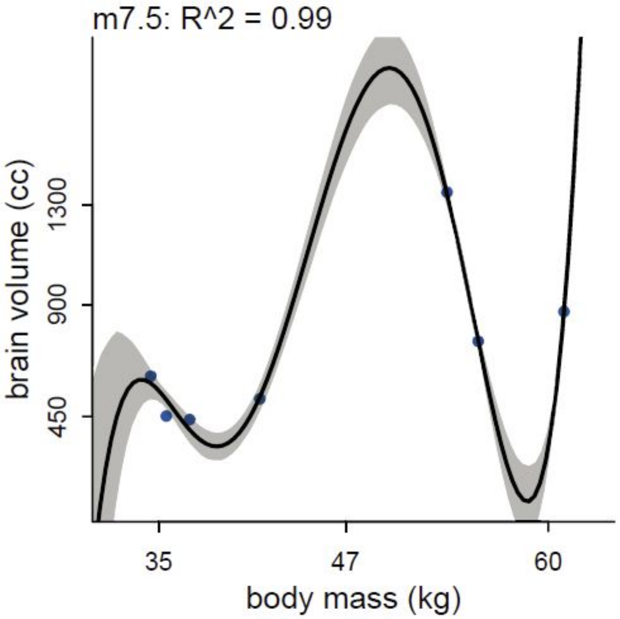
Increasing R^2



(Copy [McElreath20])

Example: Polynomial regression (degree 5)

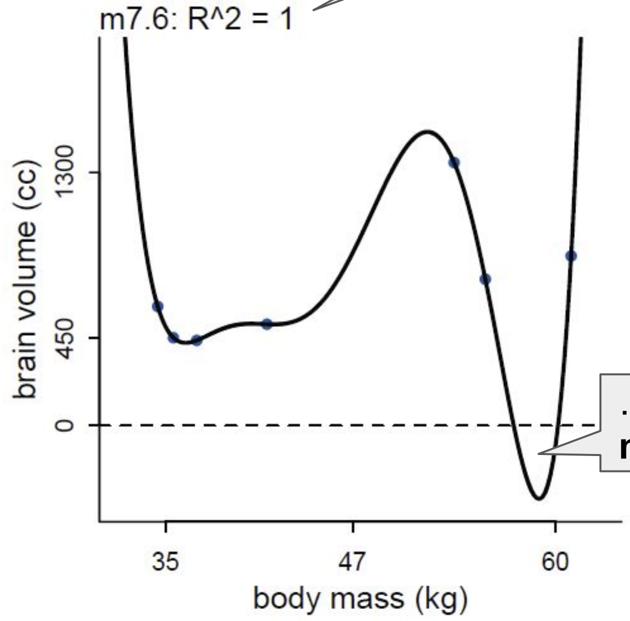
Increasing R^2



(Copy [McElreath20])

Example: Polynomial regression (degree 6)

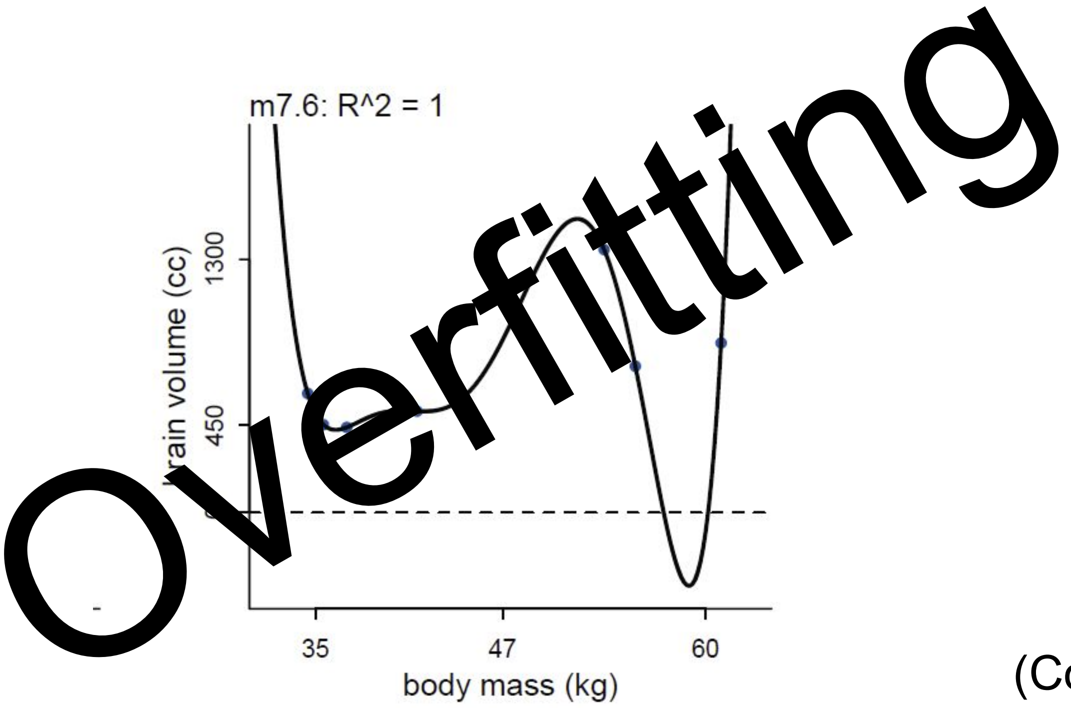
Perfect fit…



… but this model is **nonsense**.

(Copy [McElreath20])

Example: Polynomial regression (degree 6)



(Copy [McElreath20])

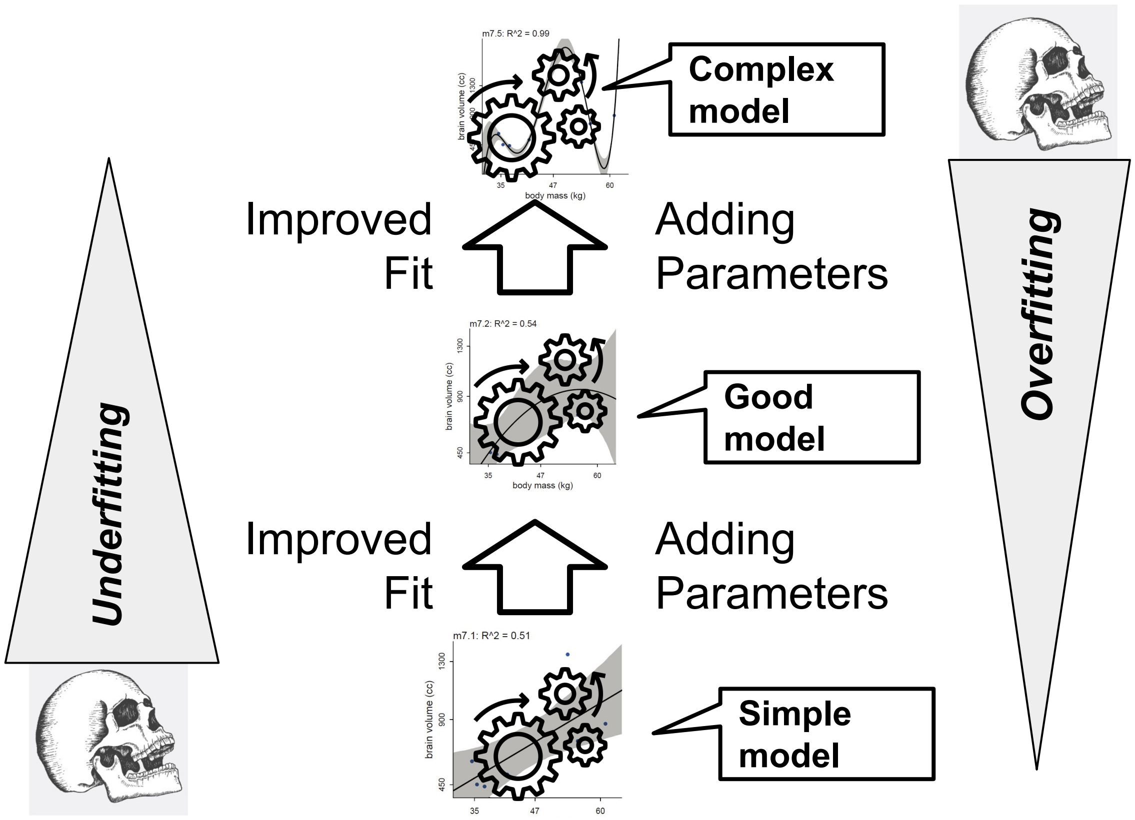
Learning too much: **overfitting**

Learning too little: **underfitting**

The balance between underfitting and overfitting is sometime called bias-variance trade-off (see [McElreath20] page 201).

Navigating between under- and over-fitting

- **Adding parameters** to a model (almost) always improves the fit, even if parameters are meaningless (**overfitting)**.
- **Not adding parameters** (and variables) to a model may learn to little (**underfitting)**.
- **Navigating between both** is the major challenge of a model comparison.



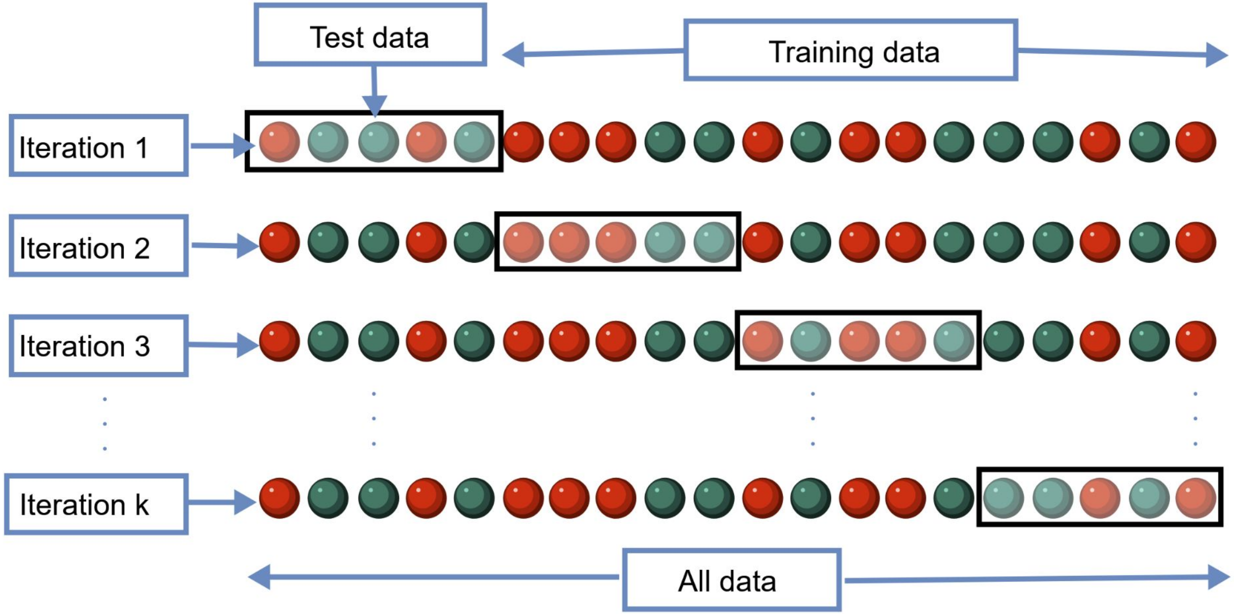
Strategies for the Navigation

Strategies for the Navigation

- Cross-validation (coming up next in-depth)
- Regularization
  - Regularization **shrink the parameters towards zero** so that the model does **not get over-excited** about the data. This can be done by priors.
- Information criteria (AIC, WAIC, PSIS)
  - **Motivation: Cross-validation is expensive** since it needs to i) split the data set several times, ii) fit a model and iii) evaluated it on the test set.
  - There are **theoretical estimates of the out-of-sample fit**, that can also be used, called **information criteria**.

Let’s go for cross-validation

K-fold cross-validation variant



Source: https://en.wikipedia.org/wiki/Cross-validation_(statistics)

Demo